

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 10269233 A

(43) Date of publication of application: 09.10.98

(51) Int. Cl

G06F 17/30

(21) Application number: 09071317

(71) Applicant: HITACHI LTD

(22) Date of filing: 25.03.97

(72) Inventor: NOJIMA HIROSHI

(54) METHOD AND DEVICE FOR DISPLAYING
RETRIEVAL RESULT OF DOCUMENT DATA
BASE

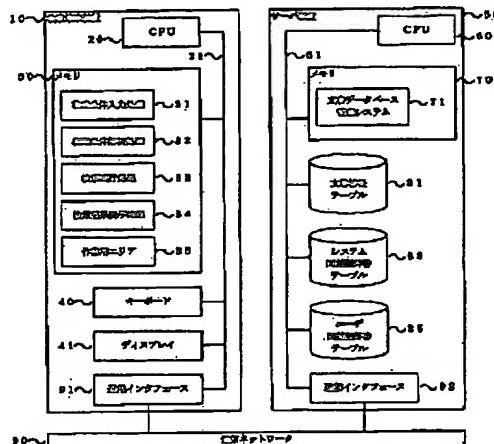
condition and can easily judge whether or not the document is necessary.

COPYRIGHT: (C)1998,JPO

(57) Abstract:

PROBLEM TO BE SOLVED: To make full-text retrieval and browsing operation by user efficient by emphasizingly displaying character strings regarding full-text conditions in a document with attributes different by kinds of full-text retrieval when a full-text retrieved document obtained from a document data base is displayed.

SOLUTION: Under retrieval conditions inputted in a process 31, object character strings of full-text retrieval are extracted in a process 32 and similar words are expanded to generate an emphasis-display character string list. For a place meeting the full-text retrieval conditions included in a document of retrieved result obtained by a process 33, matching position information is obtained in a process 34 and emphasis display attributes which are different by the kinds of full-text retrieval are set to display the document on a display 41. The user can instantaneously recognize which part of the document meets which full-text retrieval



(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-269233

(43) 公開日 平成10年(1998)10月9日

(51) Int.Cl.⁶

G 0 6 F 17/30

識別記号

F I

G 0 6 F 15/403

3 8 0 Z

3 3 0 C

審査請求 未請求 請求項の数 7 O.L. (全 24 頁)

(21) 出願番号

特願平9-71317

(22) 出願日

平成9年(1997)3月25日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 野島 博

神奈川県川崎市幸区鹿島田890番地株式会

社日立製作所情報・通信開発本部内

(74) 代理人 弁理士 小川 勝男

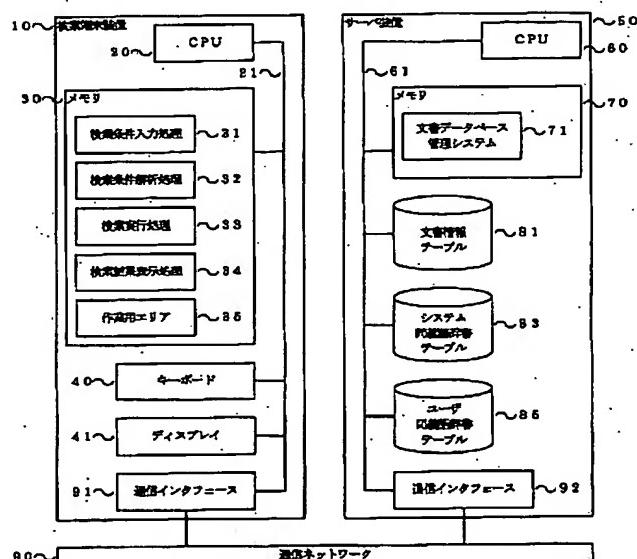
(54) 【発明の名称】 文書データベースの検索結果表示方法及び装置

(57) 【要約】

【課題】文書データベースから全文検索された文書を表示する際、文書中の全文検索条件に関わる文字列を全文検索の種類ごとに異なった属性で強調表示し、利用者の全文検索と閲覧作業の効率化を図ることである。

【解決手段】処理31で入力された検索条件より、処理32で全文検索の対象文字列を抽出し、類似語の展開も行なって、強調表示文字列リストを生成する。処理33で取得された検索結果の文書に含まれる、全文検索条件に合致した箇所に対し、処理34で合致した位置情報を取得し、全文検索の種類ごとに異なる強調表示属性を設定し、文書をディスプレイ41に表示する。利用者はその文書のどの部分がどの全文検索条件に合致したのかを即座に認識し、文書の要不要を容易に判断することができる。

図 1



【特許請求の範囲】

【請求項1】文書とその書誌的情報が格納されたデータベースから、文書に対する全文検索条件と書誌的情報に対する検索条件の組み合わせから成る検索条件式を用いて所望の文書を検索し、検索結果の文書を画面に表示する方法であって、

全文検索条件の中から全文検索の対象となる文字列を取得して、全文検索条件において類似語を含めた検索が指定されていた場合には類似語展開された結果の文字列も検索対象文字列に追加して、強調表示文字列リストを作成し、

検索結果として取得した文書データを走査し、前記強調表示文字列リストに含まれる文字列が文書中に出現する位置情報を取得して、全文検索条件と照合して強調表示箇所とその強調表示種別を決定して強調表示対象箇所リストを作成し、

検索結果文書の画面表示を行なう際に、前記強調表示対象箇所リストに基づき全文検索条件に合致した箇所に強調表示属性を設定して表示し、利用者が容易に検索結果を確認出来ることを特徴とする文書データベースの検索結果表示方法。

【請求項2】請求項1記載の全文検索条件として、指定文字列の完全一致による条件に加えて、指定文字列の同義語または異表記展開による類似語文字列の一致、二つの指定文字列とその間の距離による近傍条件一致、三つの指定文字列による範囲条件一致などの曖昧検索条件も全文検索条件として適用出来ることを特徴とする文書データベースの検索結果表示方法。

【請求項3】請求項1記載の強調表示種別として、検索結果文書中の、指定文字列の完全一致箇所、指定文字列の同義語文字列の一致箇所、指定文字列の異表記展開文字列の一致箇所、二つの指定文字列とその間の距離による近傍条件一致箇所、三つの指定文字列による範囲条件一致箇所について、それぞれ異なる形式の強調表示属性を設定することを特徴とする文書データベースの検索結果表示方法。

【請求項4】文書とその書誌的情報が格納されたデータベースから、文書に対する全文検索条件と書誌的情報に対する検索条件の組み合わせから成る検索条件式を用いて所望の文書を検索し、検索結果の文書を画面に表示する装置であって、

全文検索条件の中から全文検索の対象となる文字列を取得して、全文検索条件において類似語を含めた検索が指定されていた場合には類似語展開された結果の文字列も検索対象文字列に追加して、強調表示文字列リストを作成する手段と、

検索結果として取得した文書データを走査し、前記強調表示文字列リストに含まれる文字列が文書中に出現する位置情報を取得して、全文検索条件と照合して強調表示箇所とその強調表示種別を決定して強調表示対象箇所リ

ストを作成する手段を有し、

検索結果文書の画面表示を行なう際に、前記強調表示対象箇所リストに基づき全文検索条件に合致した箇所に強調表示属性を設定して表示し、利用者が容易に検索結果を確認出来ることを特徴とする文書データベースの検索結果表示装置。

【請求項5】請求項4記載の全文検索条件として、指定文字列の完全一致による条件に加えて、指定文字列の同義語または異表記展開による類似語文字列の一致、二つの指定文字列とその間の距離による近傍条件一致、三つの指定文字列による範囲条件一致などの曖昧検索条件も全文検索条件として適用出来ることを特徴とする文書データベースの検索結果表示装置。

【請求項6】請求項4記載の強調表示種別として、検索結果文書中の、指定文字列の完全一致箇所、指定文字列の同義語文字列の一致箇所、指定文字列の異表記展開文字列の一致箇所、二つの指定文字列とその間の距離による近傍条件一致箇所、三つの指定文字列による範囲条件一致箇所について、それぞれ異なる形式の強調表示属性を設定する手段と、

前記強調表示属性の付与された文書を画面表示する手段を有することを特徴とする文書データベースの検索結果表示装置。

【請求項7】文書の格納されたデータベースから所望の文書を検索し表示する装置を、文書データベース管理システムを稼動させるサーバ装置と、利用者が検索及び検索結果表示処理を行なうクライアント検索端末装置とを通信ネットワークで接続して構成し、全文検索を行なつて検索結果文書中の全文検索条件合致箇所を強調表示させる時に、

サーバ装置上で、多数の文書データの中から全文検索条件に合致する文書データ群を検索する手段と、

クライアント検索端末装置上で、前記検索結果の個々の文書データの中から全文検索条件に合致している箇所を検索し、その位置情報リストを構築する手段を持ち、サーバ装置とクライアント検索端末装置とで全文検索処理の負荷の分散を図ることでより効率良く検索作業を進められることを特徴とする文書データベースの検索結果表示装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、新聞記事、特許情報、裁判の判例集、社内文書などの文書情報を保存し、再利用することを目的としたデータベース管理システムに関する、特にこの文書データベースを用いて文書の検索を行なった際の効果的な検索結果表示方法及び装置に関する。

【0002】

【従来の技術】文書データベースから所望の文書を検索する場合に、文書の見出し、文書の著者、文書が作成あ

るいは公開された日付などの、文書の書誌的情報に対する検索を行なうだけでなく、文書データそのものの中に出現するキーワードに対する条件を用いて検索を行なうことが、より効果的である。

【0003】近年のデータベース管理システムでは、データベースに格納された文書データに対して、完全一致検索だけでなく、同義語検索、異表記展開検索、近傍条件検索、範囲条件検索などの、曖昧な文字列表現に対しての検索も実現可能になっている。

【0004】完全一致検索とは、指定されたキーワードそのものが文書の中に含まれているかどうかを検索条件とする検索である。

【0005】同義語検索とは、指定されたキーワードそのもの、もしくはそのキーワードの同義語が、文書の中に含まれているかどうかを検索条件とする検索である。キーワードの同義語を取得するためには、専用の同義語辞書を利用する。この同義語辞書は、文書データベース管理システムが標準で持つ一般的な同義語辞書、あるいは利用者が指定する特定の分野の同義語辞書などを、利用が必要に応じて使い分ける。これら同義語辞書の管理方式は様々であり、データベース管理システムがテーブルの一つとして管理することもある。同義語検索を行なう際、利用者はどのような同義語が検索対象になっているのかを意識する必要はなく、同義語情報取得時に参照される同義語辞書名の指定とともに、「同義語もあわせて検索するように」と指示をすれば、同義語の検索も含めた検索結果が得られる。

【0006】異表記展開検索とは、ある文書の中に指定されたキーワード、もしくはそのキーワードの異表記形、つまり異なる表記形態のものが含まれているかどうかを検索条件とする検索である。ここで異表記とは、例えば英字では、全文字が大文字のもの、全文字が小文字のもの、先頭文字のみ大文字で2文字目以降は小文字のもの、などの3通りが挙げられる。また、英数字では、同じ綴りでも2バイト文字（全角文字）のものと1バイト文字（半角文字）のものとが異表記形であると言える。異表記展開検索では、これらを区別することなく検索が行なわれる。

【0007】なお、類似語検索とは、同義語検索と異表記展開検索の総称である。

【0008】範囲条件検索とは、指定したキーワード（始点文字列）とキーワード（終点文字列）の間に、指定したキーワード（区間内文字列）が存在することを条件とした検索である。区間内文字列については前記の同義語や異表記展開を適用してもよい。文書中の複合語に関する検索、利用者の記憶が曖昧な文脈に対しての検索などに有効である。

【0009】近傍条件検索とは、指定したキーワード（基点文字列）とキーワード（端点文字列）の間の距離を条件とした検索である。端点文字列については前記の

同義語や異表記展開を適用することもでき、基点文字列と端点文字列の出現順序の指定（基点文字列が先で端点文字列が後に出現する場合に限定するか、出現順序を限定しないか）や、これらのキーワードの距離が何文字以下、何文字ちょうど、何文字以上のいずれであるかの指定も行なう。範囲条件検索と同様、文書中の複合語に関する検索、利用者の記憶が曖昧な文脈に対しての検索などに有効である。

【0010】ここで述べた完全一致検索、同義語検索、異表記展開検索、範囲条件検索、近傍条件検索では、いずれの場合も、条件に合致する箇所が含まれている位置及び個数を問わず、文書中に含まれているならば検索条件に合致したことになる。全文検索装置では、これらの検索条件を「かつ」「または」で代表される論理式を使って組み合わせることで、複合条件の指定も可能となっているのが一般的である。

【0011】このように、文書情報に対しての完全一致検索から、類似語検索、複数文字列の組み合わせによる曖昧検索へと検索技術が発展したことによって、従来行なわれてきたような、見出し情報、著者情報、日付情報といった書誌的情報に対する検索だけでなく、文書の記載内容そのものに、より直結した検索が実現出来るようになった。

【0012】しかし、検索結果は単に与えられた検索条件に合致した文書として取得するだけであり、具体的に文書のどの部分がどのような検索条件に合致したのかは検索結果情報に含まれていない。したがって、検索結果を画面に表示して利用者が閲覧する際には、完全一致検索の場合に指定されたキーワードのみを強調表示されることにとどまり、同義語検索、異表記展開検索、近傍条件検索、範囲条件検索のような、一つの検索条件の実行に複数のキーワードが係る複雑な検索では、文書中の検索条件に適合したキーワードや検索条件に適合した箇所を利用者に提示するには至っていないかった。そのため利用者は、提示された検索結果文書が所望のものかどうかを判断する際に、検索条件に適合した箇所を目で追って探し出した上で判別しなければならず、要不要の判断に手間と時間を要するといった不便を強いられていた。

【0013】

【発明が解決しようとする課題】データベース管理システムの機能の発展に伴い、文書データそのものに対する全文検索が実現出来るようになった反面、広範囲な検索によって、利用者にとって不要な文書までもが検索結果に含まれてしまうケースが多くなった。こうした冗長な検索結果については、最終的には利用者自身で文書に目を通して取捨選択を行ない、真に必要な文書かどうか判断しなければならず、この作業に要する手間と時間の低減が求められている。

【0014】全文検索利用者による閲覧過程において、検索結果文書のどの箇所がどのような全文検索条件に合

致したのかが明白になるよう画面に表示することで、利用者はその文書が自身の関心に関連のある文書なのか、関連のない文書なのかを短時間で認識することができ、関連のない文書の時にはそれを即座に読み飛ばすことで、効率良く文書の検索を進められる。

【0015】本発明の目的は、文書そのものに対して様々な全文検索条件で検索を行なった際に、検索条件に合致した箇所を効果的に強調表示し、どの部分が検索条件に合致したのかを利用者に明白に提示し、文書の取捨選択の労力を低減させる検索結果表示方法及び装置を提供することにある。

【0016】

【課題を解決するための手段】利用者は、文書データベースに登録された文書情報に対して、文書の見出し情報、著者情報、日付情報などの書誌的情報に対する検索条件、及び文書そのものに対する全文検索条件を指定し、これらを「かつ」「または」で結合する検索論理式を設定する。全文検索条件の例としては、前述したような、完全一致検索、同義語検索、異表記展開検索、近傍条件検索、範囲条件検索などがある。

【0017】利用者からの検索実行の指示により、入力された検索条件と検索論理式から検索条件式が構築される。この検索条件式に基づいて、文書データベース管理システムでは検索が実行される。また、閲覧時の強調表示対象を決定する処理で使用するため、ここでは全文検索条件に含まれる検索文字列と、検索文字列で類似語展開指定されたものについて展開した結果を取得する。この強調表示対象取得処理によって、強調表示文字列リストが生成される。

【0018】検索が実行された結果、条件に合致した文書の件数が妥当であるならば、利用者は検索結果の閲覧を開始し、一文書づつ表示させて所望の文書かどうかを確認し、所望の文書ならばファイルへの保存、印刷などの作業をする。条件に合致した文書件数が多過ぎる時には、検索条件と検索論理式を設定し直し、再度検索を実行する。

【0019】検索結果の文書を提示する際に、文書データベース管理システムから取得された文書に関して、先に生成された強調表示文字列リストに含まれる文字列の出現位置を調べ、近傍条件検索及び範囲条件検索が指定されていた時には文字列間の距離及び位置関係を調べ、条件に適合する箇所の位置情報を格納された強調表示対象箇所リストを生成する。強調表示対象箇所リストに基づき、文書に強調表示属性を設定する。

【0020】強調表示属性の種別としては、反転表示、カラー表示、太字表示、アンダーライン表示、斜体表示などがある。完全一致検索、同義語検索、異表記展開検索、範囲条件検索、近傍条件検索などの検索条件種別のそれぞれに対し、表示装置でサポートされている強調表示属性を別々に割り当てることで、検索条件に合致した

箇所を検索条件の種別情報とともに際立たせて表示する。表示属性の割り当ては、表示装置における表示可能属性の制約及び利用者の好みに応じて変更が可能である。

【0021】このような文書中の全文検索条件合致箇所の強調表示により、利用者はたくさんの検索結果文書の中から必要な文書のみを取捨選択するのが容易になり、検索作業の効率化が実現する。

【0022】

【発明の実施の形態】以下、本発明の実施の形態を詳細に説明する。

【0023】図1は、本発明の一実施例のシステム構成図を示す。実施例のシステムはクライアントサーバ型のシステムであり、通信ネットワーク90でネットワーク接続されたクライアント装置である検索端末装置10とサーバ装置50で構成される。検索端末装置10は、全体の制御及び種々の処理を実行するCPU20、CPU20で実行されるプログラム群及び処理途中のデータなどを格納するメモリ30、コマンドやデータなどを入力するためのキーボード40、検索ウィンドウを表示するディスプレイ41、及び通信ネットワーク90を介してサーバ装置50との通信を行なう通信インターフェース91からなり、各部はデータ／制御バス21を介して結ばれている。メモリ30には、磁気ディスクなどを用いた仮想メモリも含まれる。また、キーボード40には、キーボード入力を補助するためのマウスなどの入力機器も含まれる。サーバ装置50は、全体の制御及び種々の処理を実行するCPU60、CPU60で実行されるプログラム群及び処理途中のデータなどを格納するメモリ70、文書情報テーブル81、システム同義語辞書テーブル83、ユーザ同義語辞書テーブル85、及び通信ネットワーク90を介して検索端末装置10との通信を行なう通信インターフェース92からなり、各部はデータ／制御バス61を介して結ばれている。ここで、CPUとして、メインフレームで言うところの中央処理装置のみではなく、ワークステーションやパーソナルコンピュータで用いられるマイクロプロセッサやマイコンチップなどを採用しても本発明の実施には差し支えない。なお、サーバ機能とクライアント機能を通信ネットワークを介さずに同一装置上で動作するよう実装しても、本発明の実施には差し支えない。その場合はCPU20と60、データ／制御バス21と61、メモリ30と70には同じものが使われることになり、通信ネットワーク90及び通信インターフェース91と92は不要である。

【0024】メモリ30内の31～34は、特に本発明に係る処理プログラム群であり、各々、検索条件入力処理、検索条件解析処理、検索実行処理、検索結果表示処理などを実施する。これらの処理については後で詳述する。また、メモリ30内の35は作業用エリアであり、後述する検索結果文書、強調表示文字列リスト、強調表

示対象箇所リストなどを格納するために使用される。

【0025】メモリ70内の71は文書データベース管理システムであり、検索端末装置10からの検索要求を受け文書情報テーブル81に対しての検索を実行する。文書情報テーブル81には、文書データの他、文書の見出し、著者、日付などの書誌的情報も格納される。システム同義語辞書テーブル83及びユーザ同義語辞書テーブル85には同義語情報が格納され、同義語指定を伴う検索の場合に参照される。

【0026】文書データベース管理システム71では、通常のリレーションナルデータベースで取り扱われる数値や文字などの情報だけでなく、文書そのものを取り扱うことができる。文書データはそのままの形で文書格納用の属性を持ったカラムに格納される。文書格納用カラムに対しては文書専用のインデックスが定義され、インデックス中に文書のすべての文字に対する位置情報を持つことにより、文書データから目的の文字列を高速に検索することが可能である。この検索手法は全文検索と称される。全文検索条件は完全一致検索に限定されず、同義語検索、異表記展開検索といった広範囲な検索にも対応し、範囲条件検索や近傍条件検索のような複雑な検索条件にも対応する。

【0027】図2は文書情報テーブル81の構成例である。文書情報テーブル81は、文書データ及びその書誌的情報をユニークなIDで管理するテーブルである。ここでは新聞記事情報を格納した場合を例に挙げた。図2において、一つのレコード820は一つの文書（この例では記事）に対応し、レコードは文字型（あるいは数値型）の文書IDカラム810、日付型の日付カラム811、文字型の見出しカラム812、文書型の本文カラム813からなる。実施例では4カラム構成のテーブルとしたが、より多くの書誌的情報を保持するためにカラム数を増やしてテーブルを構成しても差し支えない。

【0028】図3はシステム同義語辞書テーブル83の構成例である。システム辞書テーブル83はあらかじめ作成された汎用的なものであり、一般的に利用者が新たに同義語データを追加・削除したりすることはできない。図3において、一つのレコード840は一つの同義語文字列に対応し、レコードは数値型のグループIDカラム830、文字型の同義語カラム831からなる。同一のグループIDを持つものは一つの同義語群を成し、例えばレコード840、841、842、843は同じグループID、“0001”を持つことから、“パソコン”“コンピュータ”“電子計算機”“Computer”は同じ同義語群に属することになる。同義語情報の取得手順例としては、例えば“コンピュータ”という言葉の同義語を調べる場合、まず同義語辞書テーブルの同義語カラムから“コンピュータ”を検索してそのグループID“0001”を取得し、次にグループIDに“0001”を持つものを検索し、検索結果を同義語群とし

て取得する、といった手順になる。

【0029】図4はユーザ同義語辞書テーブル85の構成例である。ユーザ同義語辞書テーブル85は、特定分野に特化した同義語情報を格納したもので、更新権限を持つ利用者によって同義語データの追加・削除が可能である。図4において、一つのレコード860は一つの同義語文字列に対応し、レコードは数値型のグループIDカラム850、文字型の同義語カラム851からなる。同一のグループIDを持つものは一つの同義語群を成し、例えば、レコード860、861、862、863は同じグループID“0001”を持つことから、“ネットワーク”“LAN”“WAN”“ISDN”は同じ同義語群に属することになる。ユーザ同義語辞書テーブルからの同義語の取得手順は、システム同義語辞書テーブルの場合と同様である。

【0030】なお、本実施例ではシステム同義語辞書テーブルもユーザ同義語辞書テーブルもグループIDによって同義語を管理する方式としたが、グループIDを用いずに、レコード間にリンクを張ることで同義語群を管理する方法でも同義語の管理は可能である。この場合にはリンク先をたどることで同義語を調べられる。本発明において、どのような同義語辞書管理方式を採用しても、発明の実施には差し支えない。

【0031】次に、図1中の検索条件入力処理31、検索条件解析処理32、検索実行処理33、検索結果表示処理34について説明する。これらは一連の検索・表示処理を構成する処理単位であり、ほぼこの順序で処理が実行され、文書の検索と検索結果の表示を行なう。全体的な処理の流れを図5のフローチャートに示す。全体の処理の流れについて大まかに説明した後、31～34の各処理の詳細を説明する。

【0032】<検索・表示処理の大まかな流れ>利用者にとって使いやすい検索・閲覧環境としては、GUI（グラフィカル・ユーザ・インターフェース）環境が挙げられる。本実施例では、図6に示すような検索ウインドウ410をディスプレイ41に表示して、利用者はキーボード40を使って対話的に検索・閲覧の指示を出して検索作業を行なう。

【0033】検索端末装置10のCPU20は、データベースに格納された文書の検索作業を実行出来るよう通信ネットワーク90を介してサーバ装置50のCPU60に検索要求に応じられるよう準備を整えさせ、検索端末装置10のディスプレイ41に検索ウインドウ410を表示して利用者に検索作業が開始出来ることを示す。図5のステップ3000の初期状態は、これを示している。ここで、サーバ装置50上では文書データベース管理システム71が既に稼動しているものとする。

【0034】ステップ3001の検索条件入力処理で、検索ウインドウ410上の4121から4124の検索条件記述部及び4125の検索論理式入力部への入力が

行なわれる。

【0035】検索実行ボタン4102の押下があるとステップ3003へ進み、検索実行ボタン4102の押下があるまでは検索条件入力処理3001へ戻り何度でも検索条件の入力・修正が行なえる（ステップ3002）。

【0036】ステップ3003の検索条件解析処理で、サーバ装置50で稼動している文書データベース管理システム71の検索実行に適用される検索条件式の構築と、検索結果を表示する際に強調表示属性を設定するために用いられる強調表示文字列リストの生成が行なわれる。続くステップ3004の検索実行処理で、前ステップで構築された検索条件式を用いてサーバ装置50のデータベース管理システム71で検索が実行され、検索結果は通信ネットワーク90を介して検索端末装置10に返される。ヒット件数表示部4134に検索で見つかった文書の数が表示され、検索結果見出し一覧表示部4131には見出しの一覧が表示される。

【0037】ヒット件数表示部4134に表示された検索結果の件数によって、利用者は再度検索条件入力をやり直すか、本文の閲覧に移るかを決定する（ステップ3005）。検索結果件数が多く検索結果見出し一覧部4131にすべての見出しを表示し切れない場合には、スクロールボタン4132及び4133によって検索結果見出し一覧表示部4131の表示をスクロールさせて検索結果の見出しを一通り見ることができる。閲覧処理に進む場合には本文表示ボタン4103を押下してステップ3006に進み、検索結果の絞り込みなどのために検索条件を設定し直す場合にはステップ3001に戻る。

【0038】ステップ3006では、選択された文書が検索結果本文表示部4141に表示される。この時、文書中の全文検索条件に合致した箇所には強調表示属性が設定され、検索結果本文表示部4141ではその箇所が強調表示される。文書全体が検索結果本文表示部4141に表示し切れない場合には、スクロールボタン4142あるいは4143の押下によって、表示文書がスクロールする。また、前一致箇所ボタン4104あるいは次一致箇所ボタン4105の押下によって、強調表示箇所間のジャンプを行なうこともできる。

【0039】閲覧を終えて次の検索を行なう時はステップ3001の検索条件入力処理へ進み、検索作業を終える時には終了ボタン4101の押下で全処理を終了する（ステップ3007）。なお、前記31～34のいずれの処理においても、終了ボタン4101の押下があった時にはその時点で全処理が終了する。

【0040】<検索条件入力処理>本実施例では、検索条件を設定する際に、大きく分けて2段階のステップがある。各カラムに対しての検索条件記述の設定と、これらの入力された検索条件の「かつ」「または」の組み合わせによる検索論理式の設定である。検索ウィンドウ4

10では、検索条件記述設定ボタン4111～4114、検索論理式設定ボタン4115、検索実行ボタン4102、検索条件記述部4121～4124、検索論理式記述部4125が本処理に関わる。なお、検索条件記述の欄を本実施例では四つ設けているが、これが四つ未満でも四つより多くても、本発明の実施においては差し支えない。図7に検索条件入力処理のフローチャートを示す。

【0041】まず、ステップ3100で処理に必要な作業用メモリの確保などの初期化処理が行なわれる。

【0042】続いて、検索ウィンドウ410上のボタンの入力を待つ。検索条件入力処理では、基本的に、ボタンを押してから条件などの入力をするという手順を踏む。ボタンが押されるまでは処理が進まない（ステップ3101）。

【0043】検索条件記述ボタン4111～4114のいずれかが押下された時は（ステップ3102）、検索条件記述入力処理で条件を対話的に入力し（ステップ3103）、押されたボタンの右に位置する、それぞれに対応する検索条件記述部（4121～4124のいずれか）に条件が表示される。

【0044】検索論理式ボタン4115が押下された時は（ステップ3104）、検索論理式入力処理で検索論理式を入力し（ステップ3105）、検索論理式表示部4125に論理式が表示される。

【0045】検索実行ボタン4102が押下された時は（ステップ3106）、検索条件解析処理へと処理が進む。これは、図5のステップ3002に相当する。

【0046】検索条件記述部入力処理のフローチャートを図8に示す。ここでは、文書情報テーブル81（図2）のどのカラムに対してどのような検索条件をつけるかを、対話的に入力する。ここで検索条件をつけられるのは、日付カラム（図2の811）、見出しカラム（図2の812）、本文カラム（図2の813）である。文書型である本文カラムに対しては全文検索条件をつけられる。

【0047】前記の図7中の検索条件記述部入力処理のフローチャートを図8に示す。まずステップ3110で、検索条件を設定する対象カラムを選択する。一つの検索条件記述では、一つのカラムに対する検索条件を設定出来る。ここでは「日付」「見出し」「本文」「クリア」の中から選択をする。

【0048】「クリア」が選択された場合（ステップ3111）、これまで設定されていた検索条件がクリアされる（ステップ3112）。

【0049】「見出し」が選択された場合（ステップ3113）、ステップ3114で見出しに含まれるキーワードの入力が行なわれる。キーワードが入力されていなかったり、入力された文字列の中に制御文字などの不適切な文字が含まれていたりしないかを調べ（ステップ3

115)、もし入力が不適切であった場合には再びステップ3114でキーワードの入力が行なわれ、適切であった場合にはステップ3120で検索条件として設定される。

【0050】「日付」が選択された場合（ステップ3116）、ステップ3117で日付範囲の設定を行なう。デフォルトでは文書データベースに格納されている文書の日付情報のうち最古のものと最新のものが設定されている。これによって、利用者は、どの程度の範囲にすれば検索にかかる時間がどの程度低減されるか見当をつけることもできる。利用者にとってデフォルトの範囲が妥当であれば、特に範囲を狭める必要はない。日付範囲を絞り込みたい場合には、開始年月日と終了年月日の指定を行なう。利用者が日付範囲を指定した場合、その日付の設定が妥当であるかどうかのチェックが行なわれる（ステップ3118）。具体的には、「13月1日」あるいは「4月31日」などといった、現実には存在しない架空の年月日が設定されていないか、終了年月日が開始年月日よりも前の日付になっていないかなどがチェックされる。日付チェックの結果、入力が不適切であった場合には再びステップ3117で日付範囲の設定が行なわれ、適切であった場合にはステップ3120で検索条件として設定される。

【0051】「本文」が選択された場合、処理はステップ3119に進み、全文検索条件入力処理が行なわれる。全文検索条件入力処理の詳細を図9のフローチャートに示す。

【0052】全文検索にはいくつかの検索種別があり、必要に応じて使い分けられる。検索種別は検索の対象となる文字列等を入力するのに先立って選択する。検索種別は指定する文字列の個数によって「簡易条件検索」「近傍条件検索」「範囲条件検索」の三つに大別できる。本実施例ではこの三つの検索条件種別を取り上げる。ステップ3130で、検索種別を対話的に選択する。

【0053】なお「簡易条件検索」とは、出現するキーワードを一つだけ指定する検索のことと、そのキーワードに対して同義語も含めた検索、異表記展開形を含めた検索といったオプション指定も可能である。両方とも指定した時には先に同義語群に展開され、それぞれの同義語に対して異表記展開を施したもの検索対象とするようによく解釈される。

【0054】「簡易条件検索」が選択された場合（ステップ3131）は、ステップ3132で、検索対象文字列の入力、同義語・異表記展開を行なうかどうかの指定を対話的に行なう。なお、同義語を指定した場合、システム同義語辞書テーブル83を用いるか、ユーザ同義語辞書テーブル85を用いるかの指定も行なわれる。入力した検索対象文字列が空であったり、入力された文字列の中に制御文字などの不適切な文字が含まれていたりし

ないかを調べ（ステップ3133）、検索条件が不適切であった時には再びステップ3132に戻る。適切であった場合には図8のステップ3120で検索条件として設定される。

【0055】「近傍条件検索」が選択された場合（ステップ3134）は、ステップ3135で、検索の基点となる基点文字列、検索の対象となる端点文字列、これらの文字列間の距離及び出現順序指定を対話的に入力する。端点文字列に関しては同義語・異表記展開を行なうかどうかの指定も行なう。基点文字列に同義語・異表記展開は適用できない。距離条件では二つのキーワードの距離が何文字以内、何文字ちょうど、あるいは何文字以上といった指定も行なう。入力された検索対象文字列のどちらかが空であったり、入力された文字列の中に制御文字などの不適切な文字が含まれたりしていないか、距離指定の値が正の数になっているかを調べ（ステップ3136）、検索条件が不適切であった時には再びステップ3135に戻る。適切であった場合には図8のステップ3120で検索条件として設定される。

【0056】「範囲条件検索」が選択された場合は、ステップ3137に進み、検索範囲の始点となる始点文字列、検索範囲の終点となる終点文字列、始点文字列と終点文字列の間に存在する区間内文字列の三つを対話的に入力する。区間内文字列に関しては同義語・異表記展開を行なうかどうかの指定も行なう。始点文字列と終点文字列には同義語・異表記展開は適用できない。入力された文字列が空であったり、入力された文字列の中に制御文字などの不適切な文字が含まれたりしていないかをチェックし、入力文字列が不適切であった時には再びステップ3137に戻る。適切であった場合には図8のステップ3120で検索条件として設定される。

【0057】ステップ3120で検索条件として設定される時に、2通りの検索条件の表現が生成される。一つは図6中の検索条件記述部4121～4124のうち、押下された検索条件記述入力ボタンに対応する箇所に表示される検索条件の表現である。これは、利用者が意味を即座に認識できるような、日本語での表示が適当である。例えば、「見出しに“家庭”が含まれる」「日付の範囲が1996年1月1日から1996年12月31日」「本文中に“パソコン”（システム辞書同義語展開・異表記展開）が含まれる」のような表現を用いることで、利用者は検索条件を違和感なく受け入れることができる。もう一つは、サーバ装置50の文書データベース管理システム71上で解釈可能な検索条件の表現である。SQL文形式で記述した例としては、「TITLE LIKE '%家庭%'」「DATE BETWEEN '19960101' AND '19961231'」「CONTAINS (DOCUMENT, SO UNDEX_EXP (SYNONYM (SYSTEM, 'パソコン')))」などがある。この表現形式は利用

者には隠蔽される。しかし、特別に表示する手段が用意されていても、本発明の実施において差し支えない。

【0058】次に、前記の図7中の検索論理式入力処理について詳述する。フローチャートを図10に示す。

【0059】まずステップ3140で、利用者によって検索論理式が入力される。

【0060】検索論理式は、図6の検索条件記述部4121～4124を対象にそれぞれを1、2、3、4、とし、これらを「かつ」「または」を現わす+と*でつなぐ。結合順序は「かつ」が「または」に優先し、任意の結合順序を指定する時には括弧を使う。例えば、「1*(2+3)」という入力は、「1であり、かつ、2または3のいずれか」という意味になる。

【0061】入力が済んだ後、続くステップ3141で、この入力が構文的に正しいかどうかをチェックする。具体的には、開き括弧と閉じ括弧の数が合っているか、1～4の数字が+と*で正しく結合されているかどうかのチェックが行なわれる。入力が不適切であった場合には、ステップ3140に戻って再度入力を行なう。

【0062】構文チェックが済んだ後、検索論理式中で使われている1～4の検索条件に対し、検索条件記述が設定済みかどうかをチェックする（ステップ3142）。検索論理式中に、検索条件記述が未設定の番号が使われていた場合には、ステップ3140に戻って再度入力を行なう。

【0063】適切な検索論理式が入力されていた場合には、ステップ3143で検索論理式としての設定が行なわれ、図6中の検索論理式表示部3125に設定した検索論理式が表示されるとともに、サーバ装置50の文書データベース管理システムで解釈可能な、「AND」

「OR」で結合されたSQLの問い合わせ文が生成される。このSQLの問い合わせ文は、一般的の利用者には隠蔽される。特別に表示する手段が用意されていても本発明の実施においては差し支えない。

【0064】<検索条件解析処理>図11に、検索条件解析処理のフローチャートを示す。この処理における目的は検索結果の本文を検索ウィンドウ410の検索結果本文表示部4141に表示する際に、本文のどの部分を強調表示すればよいか調べるのに用いる、強調表示文字列リストを生成することである。

【0065】まず、ステップ3201で、検索条件入力処理で入力された検索条件のうち、全文検索に関わる条件中に現れる文字列を抽出して取得する。具体的には、本文カラムに対する検索条件のSQL記述の中からシングルクオートされた文字列を抽出する。その際、同義語・異表記展開指定がついているかどうかを調べる。

【0066】続くステップ3202で、指定された文字列について同義語群を取得する。同義語情報はサーバ装置50の文書データベース管理システム71に問い合わせることで、システム同義語辞書テーブル81またはユ

ーザ同義語辞書テーブル83を参照して取得する。検索端末装置10で、サーバ装置50と同一の同義語辞書情報を持っている場合には、検索端末装置10内で同義語情報の取得を行なっても、本発明の実施においては差し支えない。

【0067】ステップ3203で、指定された文字列について異表記展開を実行する。異表記展開アルゴリズムは、サーバ装置50の文書データベース管理システム71で用いられているのと同じ規則を用いる。展開処理は検索端末装置10内で実行するが、サーバ装置50の文書データベース管理システム71に問い合わせを発行して取得しても、本発明の実施においては差し支えない。なお、本実施例では英字の大文字小文字の異表記展開規則のみ考慮しており、英数字の全角と半角の違いによる異表記は考慮していない。

【0068】ステップ3204で、図1の作業用エリア35内に、本処理で取得された文字列に関する、強調表示文字列リスト351の構築を行なう。

【0069】図12に「本文中に“家庭”が含まれる」と「本文中に“インターネット”から30文字以内（順不同）に“パソコン”（システム同義語展開・異表記展開）が含まれる」という二つの検索条件を「または」で結合した全文検索条件の時に構築される、強調表示文字列リスト351の一構成例を示す。なお、これ以降では、この全文検索条件を適用した検索結果表示について説明する。図11のステップ3201でリスト要素3511～3513が抽出され、ステップ3202で同義語群が加えられリスト要素3514～3516が追加され、ステップ3203で異表記展開結果3517及び3518が加えられる。リスト要素の始めの数字は検索条件記述の番号、2番目の文字列は検索対象の文字列、3番目のポインタは同義語群リンク、4番目のポインタは異表記展開リンクを現わしている。なお、強調表示対象の文字列を格納する手段にこのリスト構造と異なる構成を用いても、本発明の実施においては差し支えない。

【0070】<検索実行処理>図13に検索実行処理のフローチャートを示す。本処理において、サーバ装置50上の文書データベース管理システム71で文書検索が実行され、通信ネットワーク90を介して検索結果が検索端末装置10に返される。

【0071】まず、ステップ3301の入力バッファの初期化では、検索端末装置10のメモリ30内の作業用エリア35に置かれる、検索結果集合を保持するための領域の初期化などを実行する。

【0072】続くステップ3302で、図10のステップ3133で構築された問い合わせ文を使い、サーバ装置50上の文書データベース管理システム71に検索要求を発行する。文書データベース管理システム71はこの検索要求に応じ、必要に応じてシステム同義語辞書テーブル83あるいはユーザ定義同義語辞書テーブル85

から同義語情報を得ながら文書情報テーブル81に対し検索を実行する。インデクスをどのように用いて検索を進めるかは採用している全文検索の処理方式によって異なるが、本発明ではサーバにおける検索処理方式には特にこだわらない。

【0073】サーバ装置50で検索処理が終了すると、検索端末装置10に検索結果が返される。ステップ3303で検索結果を作業用エリア35に読み込む。検索結果が多く作業用エリア35に保持し切れない場合には、検索結果をメモリ上ではなく磁気ディスクなどに格納しても構わない。あるいは、データサイズの大きい本文カラムデータだけを後で画面に表示する段階で取得するように実装しても、本発明の実施においては差し支えない。

【0074】続くステップ3304で検索ウィンドウ410上のヒット件数表示部4134にヒットした件数を表示し、ステップ3305で検索結果の見出しの一覧を検索結果見出し一覧部4131に表示する。見出しどともに日付情報などを表示しても構わない。ヒット件数が多く、検索結果見出し一覧部4131に見出し情報を表示し切れない場合には、スクロールボタン4132もしくは4133を押下することによって、見出し一覧をスクロールさせることができる。

【0075】<検索結果表示処理>図14に検索結果表示処理のフローチャートを示す。本処理において、取得された文書データの全文検索条件に合致した箇所に強調表示属性が設定され、図6の検索結果本文表示部4141に文書が表示される。

【0076】ステップ3401の入力バッファの初期化では、検索端末装置10のメモリ30内の作業エリア35に置かれる、強調表示対象箇所リスト353の初期化などが行なわれる。

【0077】続くステップ3402で、図13のステップ3303で取得された検索結果文書のコード変換、例えば漢字コードの変換や、行末記号のコード変換などが施され、図6の検索結果本文表示部4141に表示できる形式へと変換される。

【0078】次に、この文書データに対して強調表示属性を設定する処理が実行される。

【0079】まず、ステップ3403で、強調表示文字列リスト351の文字列が文書データのどの位置に出現しているかを調べる。調査対象の文字列の一つ一つに対してKMP (Knuth-Morris-Pat) 法やBM (Boyer-Moore) 法などの検索アルゴリズムを用いて出現位置を調べ、強調表示文字列リストに出現位置を付加した、強調表示文字列位置リストを構築する。図15に示す文書データに対して、図13の強調表示文字列リスト351の文字列の出現位置を調べて追加したものを、図16の強調表示文字列位置リスト352に示す。強調表示文字列リスト351との違いは、

リスト要素内の5番目のポインタが位置情報を指すことである。強調表示対象の文字列が複数個出現している場合には、図16の位置情報リスト3542のように、一つ前の位置情報に次の位置情報がリンクされる。この格納構造を図16と異なる形式で実装しても、本発明の実施においては差し支えない。

【0080】次に、この強調表示文字列位置リスト352の中から、実際に文書を表示する際に強調表示すべき箇所を抽出する（ステップ3404）。全文検索には近傍条件検索、範囲条件検索のように複数語句の出現位置によって初めて条件に合致したと判定されるものがあり、このリストの文字列の出現位置すべてが全文検索条件に合致した箇所とは限らないからである。一つ目の条件は「本文中に“家庭”が含まれる」であり、これは単純に一致した箇所が強調表示対象になり、図16の3541と3542で示される箇所が、そのまま強調表示対象箇所となる。強調表示の範囲には“家庭”的文字数2も考慮される。二つ目の条件は「本文中に“インターネット”から30文字以内（順不同）に“パソコン”（システム同義語展開・異表記展開）が含まれる」であり、図16の3543で示される78文字目が基点となる。図16の3533からすべてのリストをたどることで、3544～3546の位置情報が候補に上がる。対象文字列が基点文字列より手前にある場合は対象文字列の末尾文字位置と基点文字列の先頭文字位置、対象文字列が基点文字列より後ろにある場合は対象文字列の先頭文字位置と基点文字列の先頭文字位置の差を比較すれば、その距離によって条件に合致しているかどうかが判別できる。3544～3546のうち、30文字以内にあるのは3546だけであり、これが強調表示対象箇所となる。図17に示す強調表示種別テーブル355に基づいて、図18に示すような強調表示対象箇所リスト356が生成される。リストの第1要素が強調表示の表示種別ID、第2要素が強調表示の始点、第3要素が強調表示の終点である。なお、強調表示対象箇所リストは利用者が変更することもできる。設定ファイルに記述する方法、もしくは図6のオプションボタン4106から対話的に設定する方法で個別に変更すればよい。

【0081】次のステップ3405で、強調表示対象箇所リストに従って文書データの対応する箇所に強調表示属性を設定する。そして、ステップ3406で、強調表示属性の設定された文書データが図6の検索結果本文表示部4141に表示される。

【0082】

【発明の効果】以上述べたように、本発明によれば、検索結果の文書情報を表示する際、文書中の検索条件に合致した箇所を強調表示することによって、利用者は当該文書が所望する文書であるかどうかを判断するのが容易になり、全文検索作業に要する時間を短縮することができる。

【図面の簡単な説明】

【図1】本発明の一実施例の全体構成を示すシステム図である。

【図2】文書情報テーブルの構成例を示す図である。

【図3】システム同義語辞書テーブルの構成例を示す図である。

【図4】ユーザ同義語辞書テーブルの構成例を示す図である。

【図5】検索・表示処理の一例の全体のフローチャートである。

【図6】検索端末装置における検索ウィンドウの一構成例を示す図である。

【図7】図5中の検索条件入力処理の一例のフローチャートである。

【図8】図7中の検索条件記述入力処理の一例のフローチャートである。

【図9】図8中の全文検索条件入力処理の一例のフローチャートである。

【図10】図7中の検索論理式入力処理の一例のフローチャートである。

【図11】図5中の検索条件解析処理の一例のフローチャートである。

【図12】強調表示文字列リストの一構成例を示す図である。

【図13】図5中の検索実行処理の一例のフローチャートである。

【図14】図5中の検索結果表示処理の一例のフローチャートである。

【図15】検索結果文書の一例である。

【図16】強調表示文字列位置リストの一構成例を示す図である。

【図17】強調表示種別テーブルの一例である。

【図18】強調表示対象箇所リストの一例である。

【符号の説明】

10 検索端末装置

20 CPU

21 データノードバス

30 メモリ

31 検索条件入力処理

32 検索条件解析処理

33 検索実行処理

34 検索結果表示処理

35 作業用エリア

40 キーボード

41 ディスプレイ

50 サーバ装置

60 CPU

61 データノードバス

70 メモリ

71 文書データベース管理システム

81 文書情報テーブル

83 システム同義語辞書テーブル

85 ユーザ同義語辞書テーブル

90 通信ネットワーク

91 通信インターフェース

92 通信インターフェース

【図2】

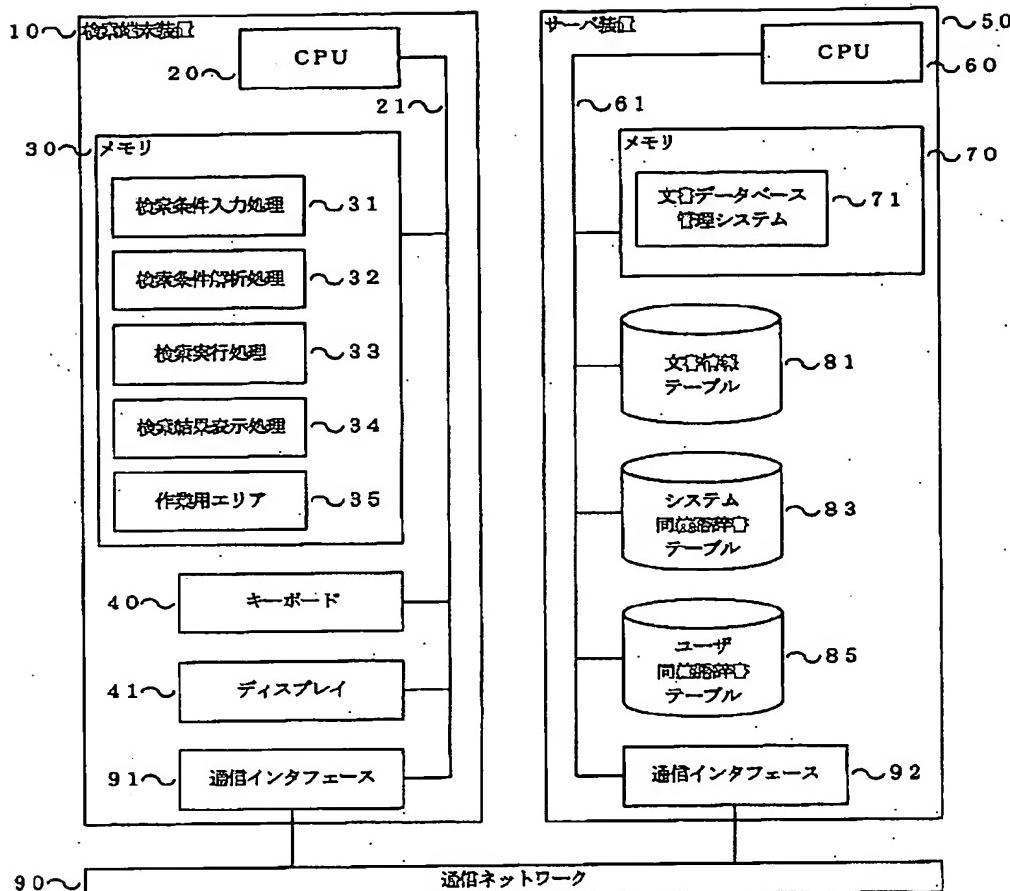
図2

文書情報テーブル81

文書ID (文字型)	日付 (日付型)	見出し (文字型)	本文 (文書型)
810 P01	811 19960914	812 “太平洋単独航海・・・”	813 “ヨットで太平洋単独横断に成功した・・・”
820～ P02	19960914	“国民所得統計速報・・・”	“経済企画庁が十三日発表した国民所得・・・”
⋮	⋮	⋮	⋮

【図1】

図1



【図15】

図15

検索結果文書352

一般家庭へのパーソナルコンピュータの普及率が急速に高まってきている。パソコンメーカー各社のし烈なシェア争いによって製品価格が下がったことに加え、昨年からのインターネットブームも手伝って、もはやパソコンは一部のマニアだけのものではなくなり、かつてのテレビやビデオのように、家庭に一台以上置かれるようになるのも、遠い先のことではなさそうだ。

~~352

【図3】

図3

システム同義語辞書テーブル8.3

	8 3 0	8 3 1
	S	S
	グループID (数値型)	同義語 (文字型)
8 4 0 ∼	0 0 0 1	パソコン
8 4 1 ∼	0 0 0 1	コンピュータ
8 4 2 ∼	0 0 0 1	電子計算機
8 4 3 ∼	0 0 0 1	C o m p u t e r
	0 0 0 2	一万円
	0 0 0 2	¥ 1.0, 000
	0 0 0 2	老万円
	0 0 0 2	老萬円
	⋮	⋮

【図18】

図18

強調表示対象箇所リスト356

表示種別 ID	始点	終点
1	3	4
1	1 3 6	1 3 7
5	7 8	8 4
10	8 5	9 6
6	9 7	1 0 1

【図4】

図4

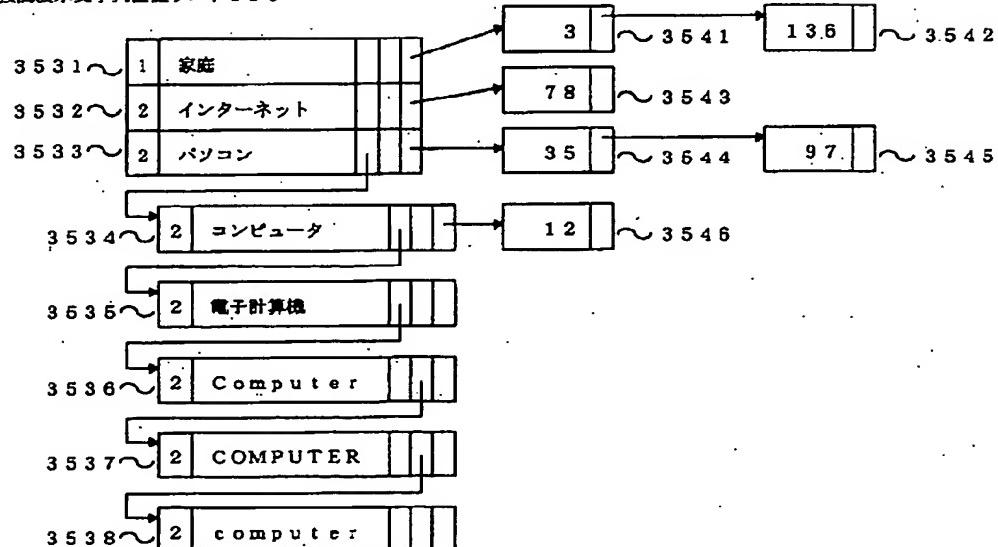
ユーザ同義語辞書テーブル85

	850	851
	グループID (数値型)	同義語 (文字型)
860～	0001	ネットワーク
861～	0001	LAN
862～	0001	WAN
863～	0001	ISDN
	0002	データベース
	0002	DBMS
	⋮	⋮
	⋮	⋮

【図16】

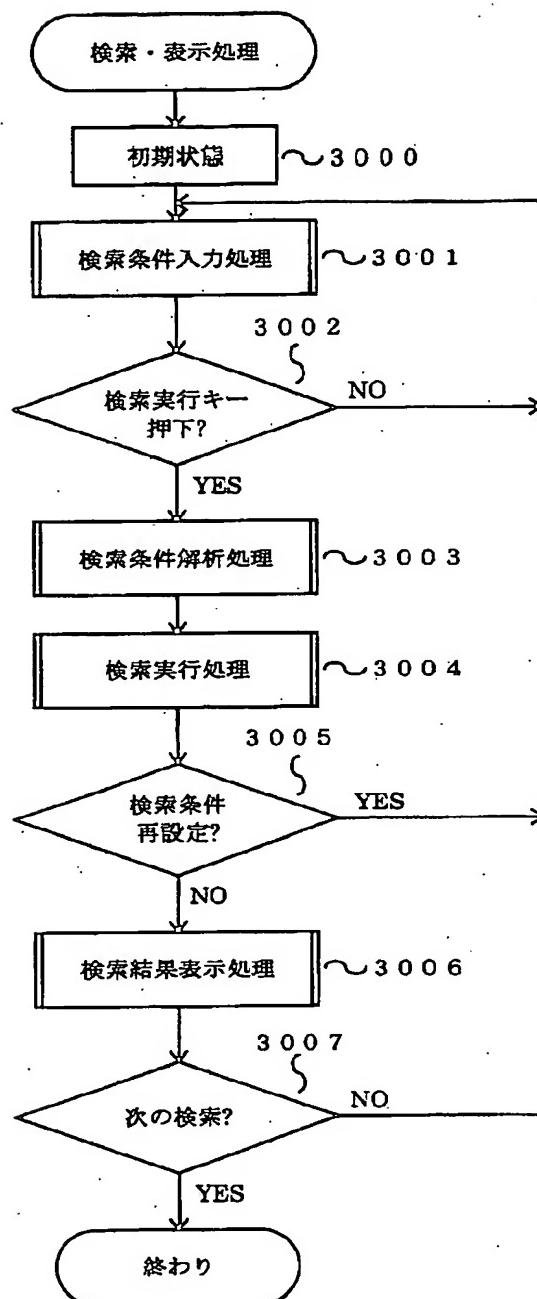
図16

強調表示文字列位置リスト353



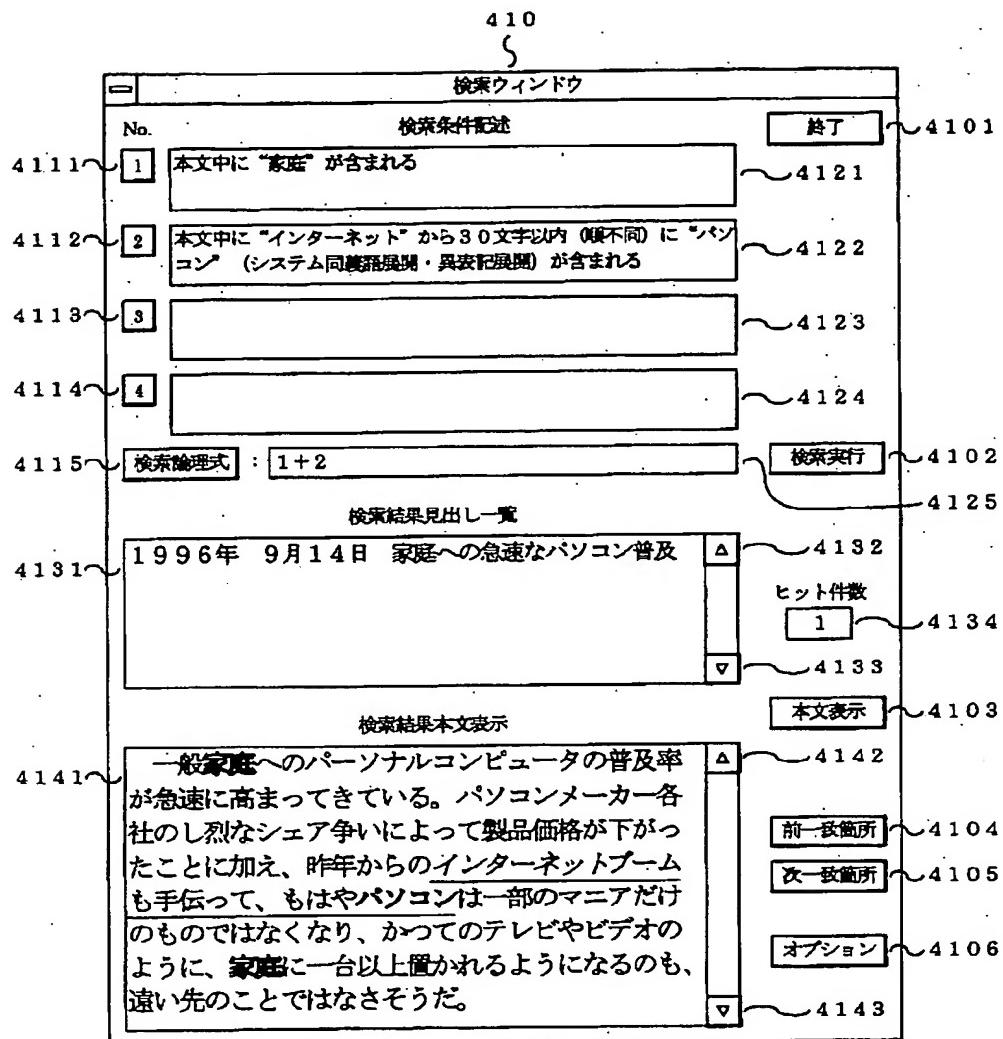
【図5】

図5



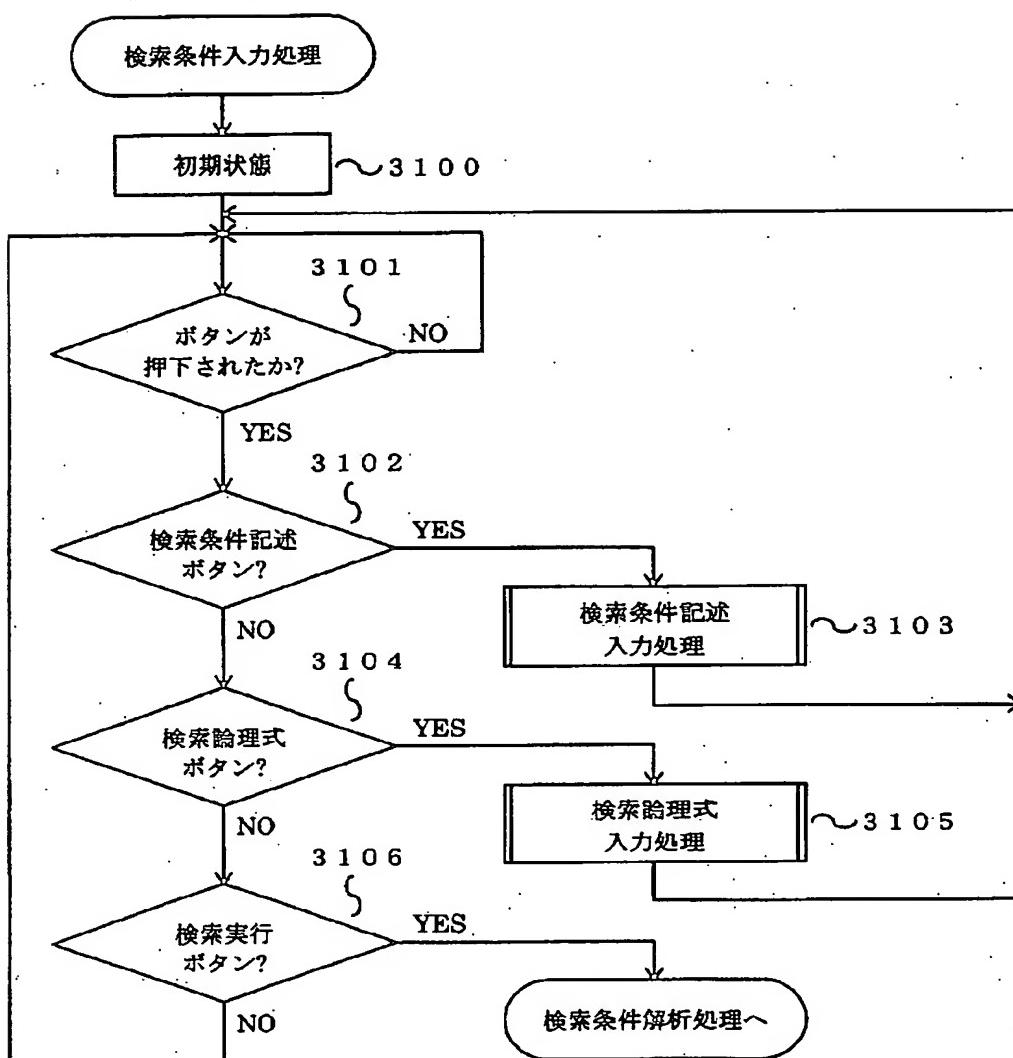
【図6】

図6



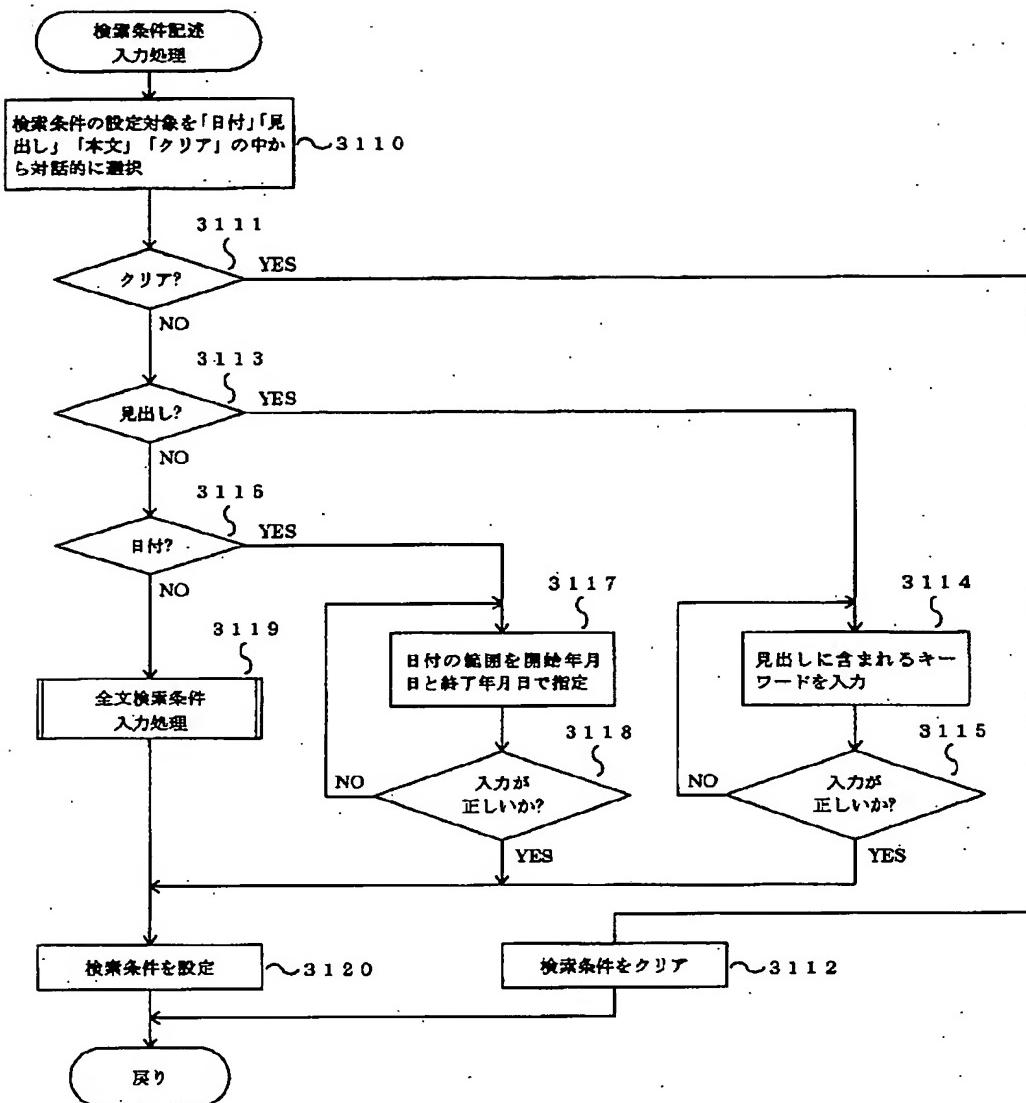
【図7】

図7



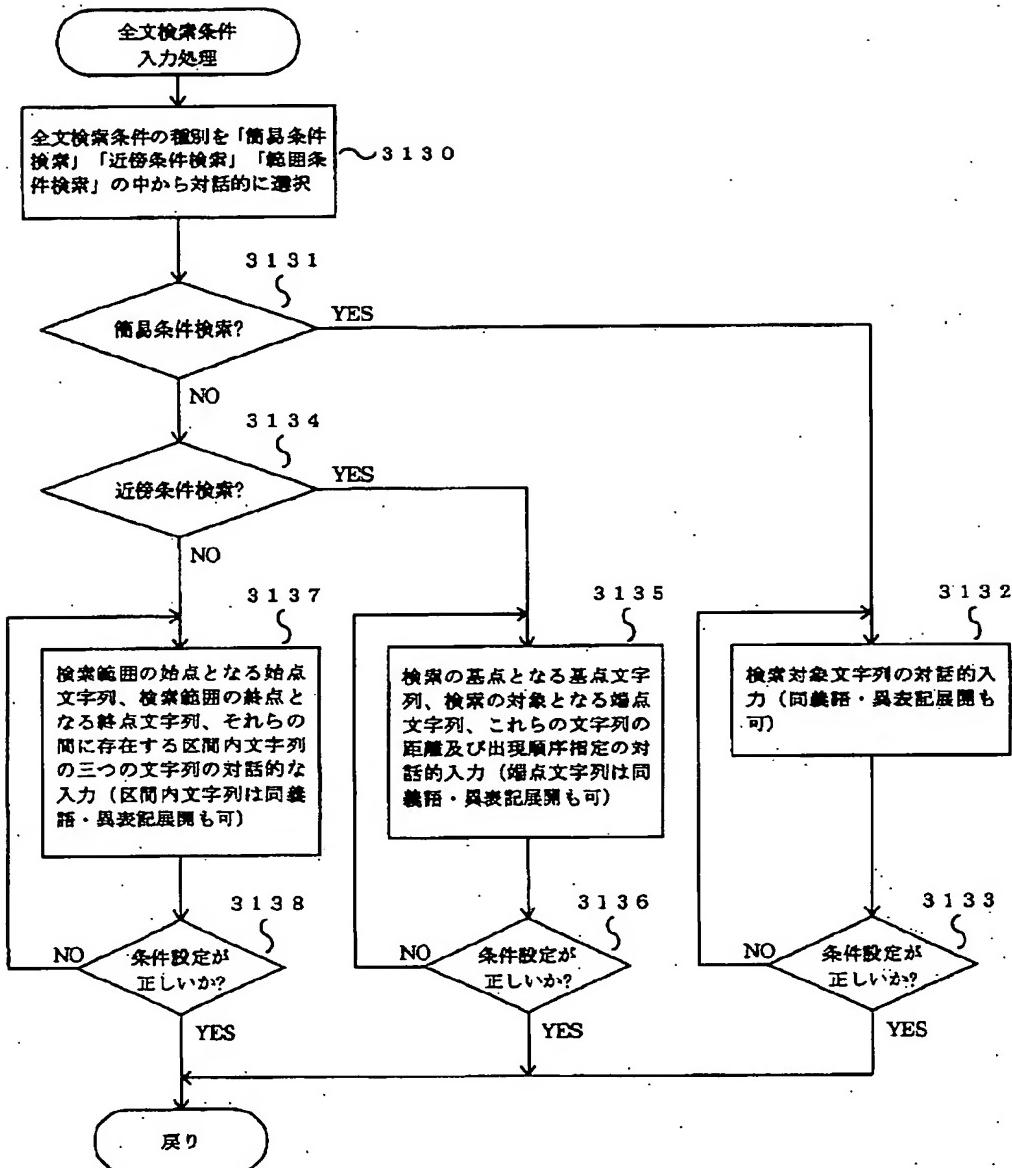
【図8】

図8



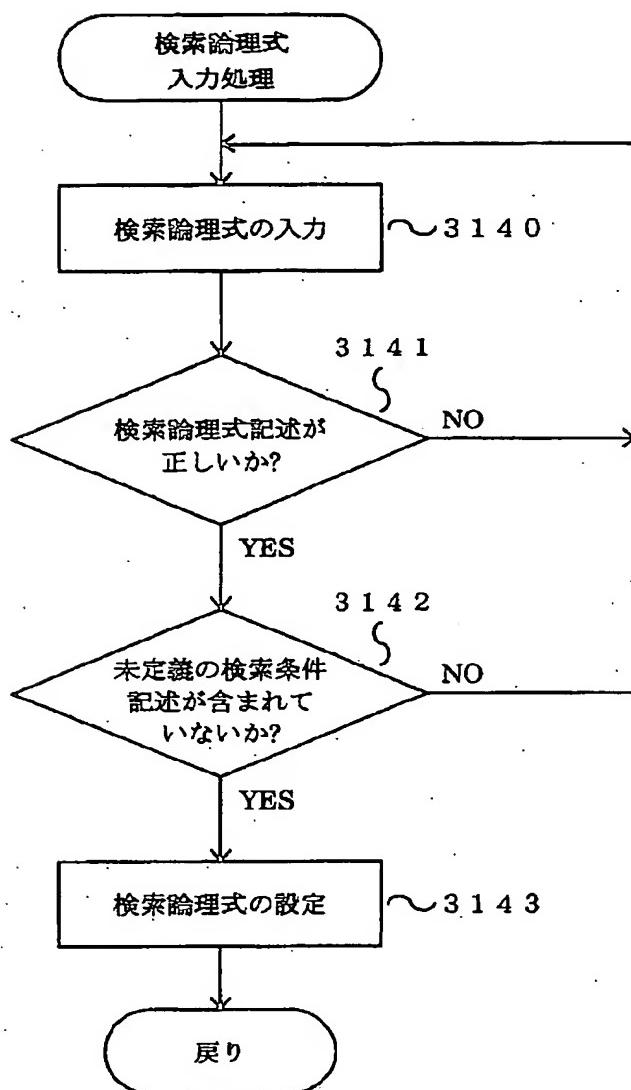
[図9]

29



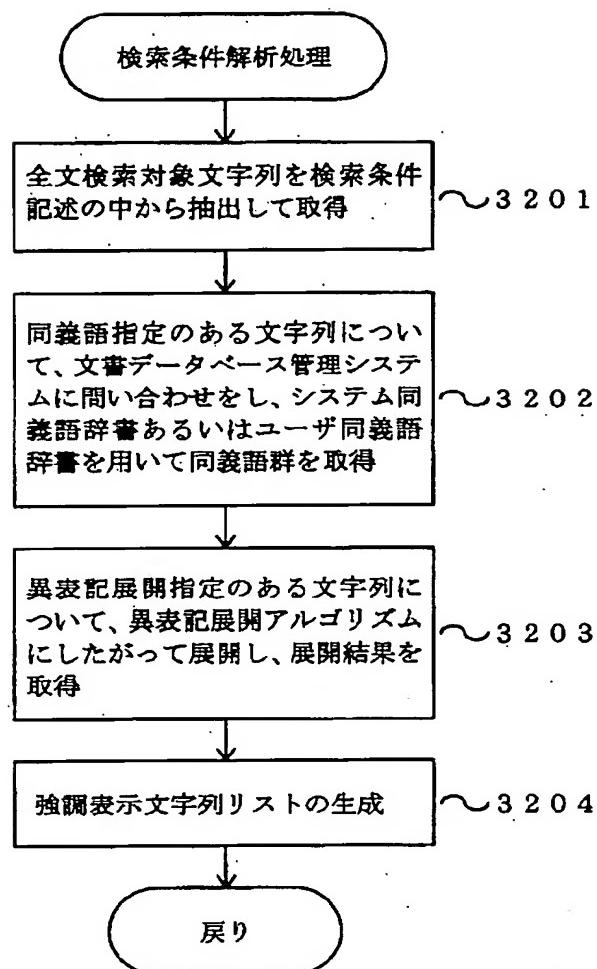
【図10】

図10



【図11】

図11



【図12】

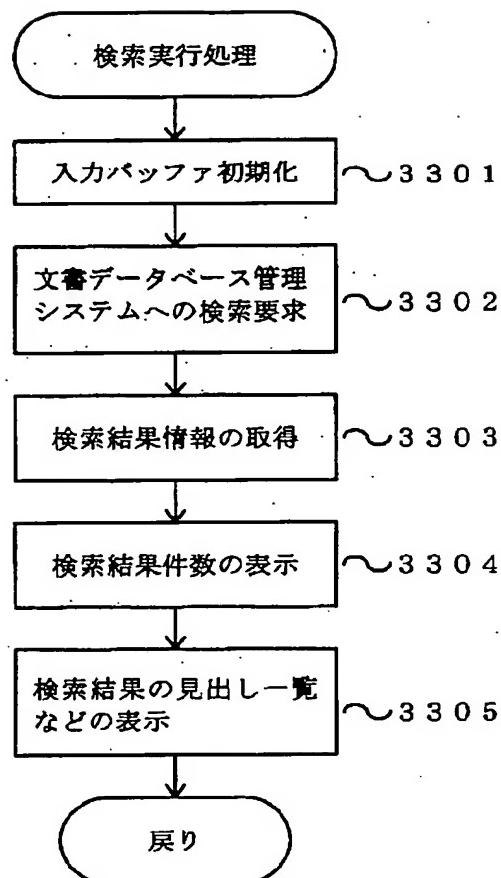
図12

強調表示文字列リスト351

3511～	1	家庭		
3512～	2	インターネット		
3513～	2	パソコン		
3514～	2	コンピュータ		
3515～	2	電子計算機		
3516～	2	C o m p u t e r		
3517～	2	C O M P U T E R		
3518～	2	c o m p u t e r		

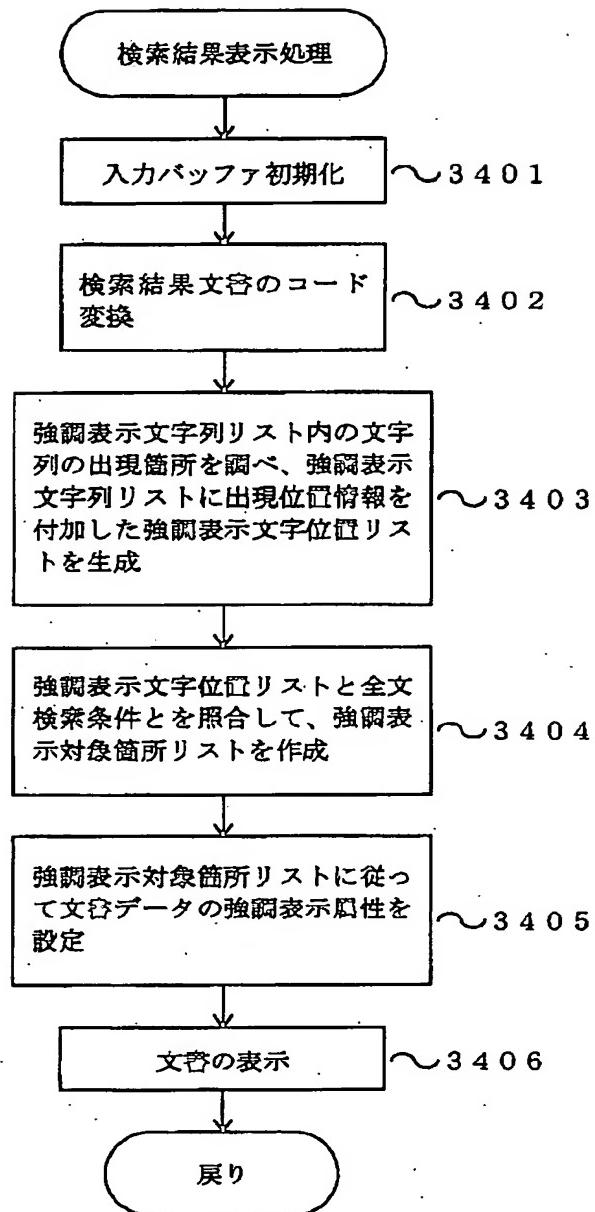
【図13】

図13



【図14】

図14



【図17】

図17

強調表示種別テーブル355

表示種別ID	意味	強調表示種別
1	簡易条件検索の完全一致文字列	太字
2	簡易条件検索の同義語文字列	太字
3	簡易条件検索の異表記展開文字列	太字
4	簡易条件検索の同義語・異表記展開文字列	太字
5	近傍条件検索の始点文字列	斜体・アンダーライン
6	近傍条件検索の始点文字列（完全一致）	太字・アンダーライン
7	近傍条件検索の始点文字列（同義語）	太字・アンダーライン
8	近傍条件検索の始点文字列（異表記展開）	太字・アンダーライン
9	近傍条件検索の始点文字列（同義語・異表記展開）	太字・アンダーライン
10	近傍条件検索の始点文字列と終点文字列の間	アンダーライン
11	範囲条件検索の始点文字列	斜体・アンダーライン
12	範囲条件検索の終点文字列	斜体・アンダーライン
13	範囲条件検索の区間文字列（完全一致）	太字・アンダーライン
14	範囲条件検索の区間文字列（同義語）	太字・アンダーライン
15	範囲条件検索の区間文字列（異表記展開）	太字・アンダーライン
16	範囲条件検索の区間文字列（同義語・異表記展開）	太字・アンダーライン
17	範囲条件検索の始点文字列と終点文字列の間	アンダーライン